

Genome sequence of *Oceanicola* sp. strain MCTG156(1a) isolated from a Scottish coastal phytoplankton net sample

Citation for published version:

Gutierrez, T, Whitman, WB, Huntemann, M, Copeland, A, Chen, A, Vargese, N, Kyrpides, NC, Pillay, M, Ivanova, N, Mikhailova, N, Mukherjee, S, Stamatis, D, Reddy, TBK, Ngan, CY, Chovatia, M, Daum, C, Shapiro, N & Woyke, T 2017, 'Genome sequence of *Oceanicola* sp. strain MCTG156(1a) isolated from a Scottish coastal phytoplankton net sample', *Genome Announcements*, vol. 5, no. 32, e00796-17.
<https://doi.org/10.1128/genomeA.00796-17>

Digital Object Identifier (DOI):

[10.1128/genomeA.00796-17](https://doi.org/10.1128/genomeA.00796-17)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Genome Announcements

Publisher Rights Statement:

© 2017 Gutierrez et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Genome sequence of *Oceanicola* sp. strain MCTG156(1a) isolated from a Scottish coastal phytoplankton net sample

Tony Gutierrez,^a William B. Whitman,^b Marcel Huntemann,^c Alex Copeland,^c Amy Chen,^c Neha Vargese,^c Nikos C. Kyrpides,^c Manoj Pillay,^c Natalia Ivanova,^c Natalia Mikhailova,^c Supratim Mukherjee,^c Dimitrios Stamatis,^c T.B.K. Reddy,^c Chew Yee Ngan,^c Mansi Chovatia,^c Chris Daum,^c Nicole Shapiro,^c Tanja Woyke^c

^a School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh,

UK;

^b Department of Microbiology, University of Georgia, Athens, Georgia, USA;

^c DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598-1698, USA.

***Correspondence:**

Dr. Tony Gutierrez

School of Engineering and Physical Sciences

Heriot-Watt University

Edinburgh EH14 4AS, U.K.

tony.gutierrez@hw.ac.uk

Running title: Genome sequence of a *Oceanicola* species

Abstract

Oceanicola sp. strain MCTG156(1a) was isolated from a phytoplankton net sample collected on the west coast of Scotland and selected based on its ability to degrade hydrocarbons. Here, we present the genome sequence of this strain, which is 3,881,122 bp with 3,949 genes and an average G+C content of 62.7%.

Oceanicola sp. strain MCTG156(1a) was isolated from a phytoplankton net sample that was trawled in 2009 at a sampling station designated LY1 located on the west coast of Scotland near Oban, Argyll. The strain was isolated by enrichment with phenanthrene in Zobell's 2216 marine medium at 10-fold dilution. Colonies on agar plates sprayed with phenanthrene produced distinct halos that indicated the strain's ability to degrade the hydrocarbon. Based on 16S rRNA gene sequence identity, the closest type species was *Oceanicola pacificus* strain W11-2B^T, which had been isolated from a pyrene-degrading consortium that was enriched from sediment from the Pacific Ocean (1).

Here, we report the genome sequence of *Oceanicola* sp. strain MCTG156(1a). Genomic DNA was sequenced through the DOE Joint Genome Institute 2014 Genomic Encyclopedia of Type Strains, Phase III study (2) using the Pacific Biosciences (PacBio) technology. A Pacbio SMRTbellTM library was constructed and sequenced on the PacBio RS platform, which generated 239,103 filtered subreads totaling 750.9 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov>. The raw reads were assembled using HGAP (version: 2.1.1) (3). The final draft assembly produced 5

scaffolds containing 5 contigs totaling 3.9 Mbp in size and input read coverage of 217.9X.

Project information is available in the Genomes OnLine Database (4). Genes were identified using Prodigal (5), as part of the JGI's microbial annotation pipeline (6). The predicted coding sequences (CDSs) were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool (7) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA (8). Other noncoding RNAs, such as the RNA components of the protein secretion complex and RNase P, were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (<http://infernal.janelia.org>). Additional analysis and manual functional annotation was performed within the Integrated Microbial Genomes–Expert Review (IMG ER) platform (<http://img.jpi.doe.gov>) developed by the Joint Genome Institute, Walnut Creek, CA, USA (9).

The complete genome sequence length was 3,881,122 bp with a G+C content of 62.7%. The genome contained 3,949 genes (3,881 protein-coding genes) with functional predictions for 3,226 of them. A total of 68 RNA genes were detected. Other genes, characteristic for the genus, are given in the IMG database (10).

Nucleotide sequence accession number. The draft genome sequence of *Oceanicola* sp. strain MCTG156(1a) obtained in this study was deposited in GenBank as part of BioProject no. PRJNA224116, with individual genome sequences submitted as whole-genome shotgun projects under the accession no. JQMY00000000.

ACKNOWLEDGMENT

T.G. was supported by a Marie Curie International Outgoing Fellowship (PIOF-GA-2008-220129) within the 7th European Community Framework Program. Partial support was also provided from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 635340.

Commented [TG1]: I need to add this to the final version when accepted.

The work was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

1. Yuan J, Lai Q, Wang B, Sun F, Liu X, Du Y, Li G, Gu L, Zheng T, Shao Z. 2009. *Oceanicola pacificus* sp. nov., isolated from a deep-sea pyrene-degrading consortium. *Int J Syst Evol Microbiol* 59:1158-1161.
2. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, De Vos P, Vandamme P, Eisen JA, Garrity G, Hugenholtz P, Kyrpides NC. 2015. Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains. *Standards Genomic Sci* 10:26.
3. Chin C, Alexander D, Marks P, Klammer A, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler E, Turner S, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563-569.
4. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezhenska O, Isbandi M, Thomas A, Ali RS, Kyrpides NC, and Reddy TBK. 2017. Genomes OnLine

99 Database (GOLD) v.6: Data updates and feature enhancements. *Nucleic Acids*
100 *Res.* 45(D1):D446-D456.

101 5. Hyatt D, Chen GL, Lacascio PF, Land ML, Larimer FW, Hauser LJ. 2010.
102 Prodigal: prokaryotic gene recognition and translation initiation site
103 identification. *BMC Bioinformatics* 11:119.

104 6. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D,
105 Palaniappan K, Szeto E, Pillay M, Chen IM-A, Pati A, Nielsen T, Markowitz
106 VM, Kyrpides NC. 2015. The Standard Operating Procedure of the DOE-JGI
107 Microbial Genome Annotation Pipeline (MGAP v.4) *Stand Genom Sci.* 2015,
108 10:86.

109 7. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of
110 transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.

111 8. Pruesse E, Quast C, Knittel, Fuchs B, Ludwig W, Peplies J, Glckner FO. 2007.
112 SILVA: a comprehensive online resource for quality checked and aligned
113 ribosomal RNA sequence data compatible with ARB. *Nuc Acids Res* 35: 2188–
114 7196.

115 9. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC.
116 2009. IMG ER: a system for microbial genome annotation expert review and
117 curation. *Bioinformatics* 25:2271–2278.

118 10. Chen IA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A,
119 Huang J, Andersen E, Huntemann M, Varghese N, Hadjithomas M, Tennessen
120 K, Nielsen T, Ivanova NN, Kyrpides NC. 2017. IMG/M: integrated genome and
121 metagenome comparative data analysis system. *Nucleic Acids Res.*
122 45(D1):D507-D516.